NONCONVEX COMPRESSIVE SENSING AND RECONSTRUCTION OF GRADIENT-SPARSE IMAGES: RANDOM VS. TOMOGRAPHIC FOURIER SAMPLING

Rick Chartrand

Los Alamos National Laboratory,
rickc@lanl.gov

ABSTRACT

Previous compressive sensing papers have considered the example of recovering an image with sparse gradient from a surprisingly small number of samples of its Fourier transform. The samples were taken along radial lines, this being equivalent to a tomographic reconstruction problem. The theory of compressive sensing, however, considers random sampling instead. We perform numerical experiments to compare the two approaches, in terms of the number of samples necessary for exact recovery, algorithmic performance, and robustness to noise. We use a nonconvex approach, this having previously been shown to allow reconstruction with fewer measurements and greater robustness to noise, as confirmed by our results here.

Index Terms— Image reconstruction, compressive sensing, nonconvex optimization.

1. INTRODUCTION

Results of Candès, Romberg, and Tao [1] and Donoho [2] demonstrated that sparse images can be reconstructed from fewer linear measurements than previously thought possible. The family of results, algorithms, and applications that have resulted is known as *compressive sensing* (among other related terms). The results, originally proved in the context of Fourier transform sampling, apply not only to images that have relatively few nonzero pixels, but those that are sparse with respect to some basis. The approach is to solve the following convex optimization problem:

$$\min_{u} \|\Psi u\|_1, \text{ subject to } \Phi x = b. \tag{1}$$

Here, x is an image, treated as a vector in $\mathbb{R}^N$, where N is the number of pixels; Ψ is a change-of-basis matrix, ideally such that Ψx is sparse; Φ is the measurement or sampling operator, represented as an $M \times N$ matrix; and $b = \Phi x$ is the M-dimensional vector of linear samples of x. Then, provided Φ and Ψ are sufficiently *incoherent* (roughly that rows of Φ can't be sparsely expressed in terms of columns of Ψ, and vice versa; see [3]), the solution to (1) will be exactly x, even when M is much smaller than N. For example, if Φ consists of randomly-chosen entries (from any of several known distributions), or measures randomly selected entries of the Fourier transform, then (1) recovers x exactly with high probability, provided

$$M \geq CK \log(N/K) \tag{2}$$

for some constant C, where K is the image sparsity $\|\Psi x\|_0$.

Numerical results in [4] showed that one can recover x from fewer linear measurements by replacing the ℓ^1 norm in (1) with the ℓ^p quasi-norm, where $0 < p < 1$:

$$\min_{u} \|\Psi u\|_p, \text{ subject to } \Phi x = b. \tag{3}$$

This gives a nonconvex optimization problem, but one that in practice appears to be solvable using simple algorithms [4, 5]. Theoretical results [6] show for the case of Gaussian measurements that (2) then generalizes to

$$M \geq C_1(p)K + pC_2(p) \log(N/K), \tag{4}$$

where $C_1(p)$ and $C_2(p)$ decrease as p approaches zero. Thus the dependence of the sufficient number of measurements M on the number of pixels N vanishes as $p \to 0$.

Also considered in [1] is the case of images having a sparse gradient. This doesn't quite fit into the above framework, as one must consider either the vector-valued image ∇x, or the scalar-valued but nonlinear function $|\nabla x|$. The assumption that x at least be well-approximated by an image with sparse gradient has been used in medical contexts [7], is useful in nondestructive evaluation of parts of piecewise-constant density (apart from defects), and can even be regarded as the reason for the success of total-variation image restoration [8]. In the case of partial Fourier measurements, the condition (2) on the number of random Fourier samples was shown in [1] to suffice with high probability for exact recovery of images having gradient sparsity at most K. This was demonstrated with the example of the Shepp-Logan phantom, a common test image from the medical imaging field. By sampling the Fourier transform along 22 radial lines, obtaining 5481 complex measurements $b = \Phi x$ of the 256×256 phantom x, the phantom is recovered exactly by solving

$$\min_{u} \|\nabla u\|_p, \text{ subject to } \Phi x = b \tag{5}$$

with $p = 1$. In [4], sampling along 18 radial lines (4505 samples) was shown to suffice for $p = 1$, but using $p = 1/2$ reduces the necessary number of lines to 10 (or 2521 measurements, less than 4% of the total Fourier transform). Using a nonconvex problem reduces the number of measurements needed by 44%. The problem (5) was solved using a simple gradient descent algorithm.

The reason for sampling the Fourier transform along radial lines is that it makes the example equivalent (up to discretization issues) to a tomographic reconstruction problem, such as typically encountered in X-ray CT imaging. By the Fourier slice theorem, a line of samples of the 2-D Fourier transform of an image u is the same as the 1-D Fourier transform of the "mass projection" of u perpendicular to that line (i.e., a slice of the Radon transform of u). Thus, sampling the Fourier transform along L radial lines is tantamount to having L X-ray images from which to reconstruct x. (This is equally true for the more useful case of 2-D X-rays of a 3-D object.)

This raises a question. Is it better to sample along radial lines, as in X-ray tomography, or to sample randomly, as in the Candès-Romberg-Tao theorem? On the one hand, in the context of compressive sensing, random sampling is a convenient device for ensuring that the measurements are (very likely to be) sufficiently incoherent. The probabilistic guarantees do not apply to structured sampling. On the other hand, it is entirely plausible that "X-ray sampling" might confer a tomographic benefit, and perform better than suggested by compressive sensing theory.

Below we present the initial results of numerical experiments addressing this question from three perspectives. The primary consideration will be a comparison of how many samples are required to successfully recover a sparse-gradient image with each sampling method. We will also compare the performance of a particular algorithm for the two approaches, and examine the effect of noise. General results in [9] for the convex ($p = 1$) case and [10] for the nonconvex ($p < 1$) case show generally that the ability of (1) and (3) to recover sparse signals is robust to approximation, whether in the sense of noise in the measurements b, or in the signal x being only approximately sparse. This does not address the issue of which form of Fourier sampling is more robust.

2. ALGORITHM

The gradient descent algorithm used in [4] converges slowly. The iteratively-reweighted least squares (IRLS) approach used in [5, 6] for solving (3) (with Ψ being the identity) was successful and converged rapidly, so we modify it for solving (5). The approach does not admit a gradient in the objective, but it does if we use the following unconstrained approximation:

$$\min_u \alpha \|\nabla u\|_p^p + \|\Phi u - b\|_2^2, \qquad (6)$$

with α being a parameter. This is now identical to the nonconvex extension of total-variation regularization considered in [11]. The algorithm used there was a straightforward extension of the fixed-point method of Vogel and Oman [12]. Consider the Euler-Lagrange equation of (6):

$$0 = -\alpha \nabla \cdot \left(|\nabla u|^{p-2} \nabla u \right) + \Phi^T (\Phi u - b), \qquad (7)$$

where for convenience we have ignored factors of $1/p$ and $1/2$. We solve (7) iteratively, by substituting the previous iterate u_{n-1} into $|\nabla u|$, then letting u_n be the solution of the resulting linear equation:

$$\left(-\alpha \nabla \cdot \left(|\nabla u_{n-1}|^{p-2} \nabla \right) + \Phi^T \Phi \right) u_n = \Phi^T b. \qquad (8)$$

Note that this is equivalent to the IRLS approach. In [5], it was shown that a crucial aspect of the success of IRLS is the mollification of $|\nabla u|$: we replace $|\nabla u_{n-1}|$ by $\sqrt{|\nabla u_{n-1}|^2 + \epsilon}$, initially for a relatively large $\epsilon = 1$, then successively decrement by factors of 10 after convergence. In [5], this strategy drastically improved the ability of an IRLS algorithm for solving (3) to recover sparse signals, in comparison with previous results for the FOCUSS algorithm [13]. Here, in addition to decrementing ϵ we simultaneously decrement α by factors of 10, which causes the solution of (6) to approach that of (5). This process is repeated until every pixel value of u is within 10^{-14} of the corresponding value of x, unless the procedure fails, in which case the process is continued until no further convergence is observed.

3. EXPERIMENTS

We begin with x being the 128×128 Shepp-Logan phantom (Fig. 1). We compare the performance of the algorithm, using $p = 1/2$, as smaller values of p do not significantly decrease the number of measurements required [5, 6], while decreasing the robustness to non-sparsity in x [10]. The smallest number of radial line samples required for exact reconstruction is 10, or 1241 total samples. For random samples, we find 721 to be sufficient, and 711 not. Thus, random sampling requires many fewer samples. In Fig. 4(a), we see that the radial sampling results in convergence to x in many fewer iterations. However, 1241 random samples gives near convergence sooner, as measured by the sup-norm error being less than $1/128$, as this results in the MATLAB output image being identical to x.

We also consider the effect of noise. It is not clear how best to add noise in the Fourier domain, so we add Gaussian noise to x instead. This gives a reasonable corruption of the Fourier measurements, and also addresses robustness to non-sparsity. We add noise of mean zero and standard deviation 0.01 (x being $[0, 1]$-valued), for an SNR of 26.5 dB. In Figs. 1 and 2, we see the results for the sampling described above, as well as for 18 radial lines, being the fewest required for reconstruction in the noiseless case when $p = 1$, and the same number (2201) of random samples. Radial sampling and a similar number of random samples give a similar SNR, but with random sampling the finer structures are preserved a little better.
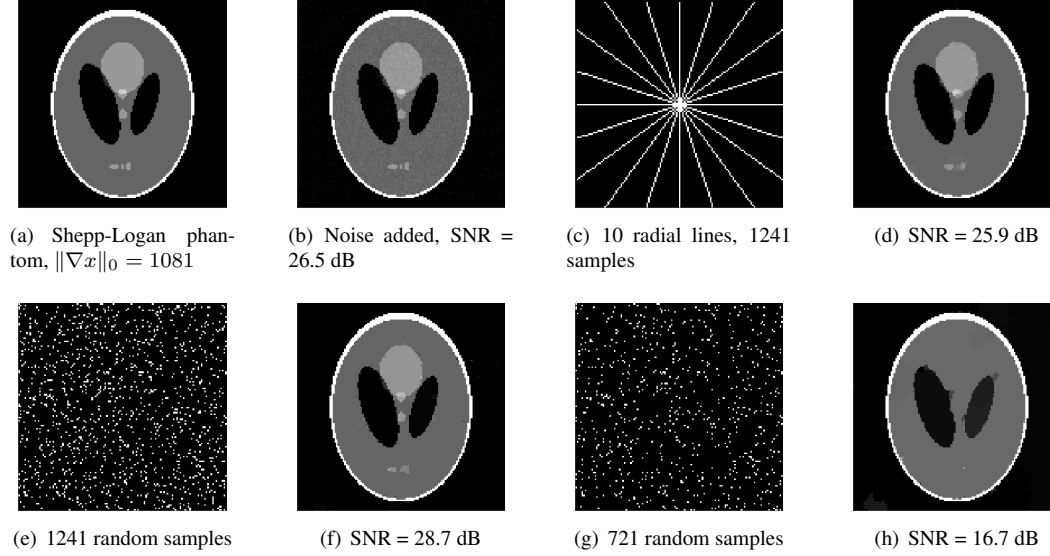
(a) Shepp-Logan phantom, $\|\nabla x\|_0 = 1081$

(b) Noise added, SNR = 26.5 dB

(c) 10 radial lines, 1241 samples

(d) SNR = 25.9 dB

(e) 1241 random samples

(f) SNR = 28.7 dB

(g) 721 random samples

(h) SNR = 16.7 dB

Fig. 1. Radial and random sampling of the Fourier transform compared for the Shepp-Logan phantom and a noisy version of it. Radial sampling does not do quite as well with the noisy phantom as the same number of random samples. Many fewer random samples are needed for the noiseless phantom, but gives much poorer results with the noisy phantom.

With a smaller number of random samples, though sufficient in the noiseless case, with noise the results are poorer. Also, using $p = 1$ gives much poorer results than $p = 1/2$, as can be expected from the fact that more measurements are needed for $p = 1$ reconstruction to be successful in the noiseless case.

We also consider an example of a very simple object, to see if the tomographic approach can be successful with as few samples as random sampling. We use an ellipse, for which 3 X-ray projections are known to be sufficient for reconstruction to be possible [14]. Results are in Figs. 3 and 4(b). We find 3 radial lines to be enough for near reconstruction, the discrepancy likely due to the coarse pixellation of the ellipse. The same number (253) of random samples is woefully insufficient. Samples along 4 radial lines are enough for perfect reconstruction, with convergence being much faster than the same number (505) of random samples. 310 random samples suffice for reconstruction, fewer than required by radial sampling for this image, but likely not for a true ellipse.

4. REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[3] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.

(a) 18 radial lines, 2201 samples

(b) $p = 1$, SNR = 19.9 dB

(c) $p = 1/2$, SNR = 33.4 dB

(d) 2201 random samples

(e) $p = 1$, SNR = 20.4 dB
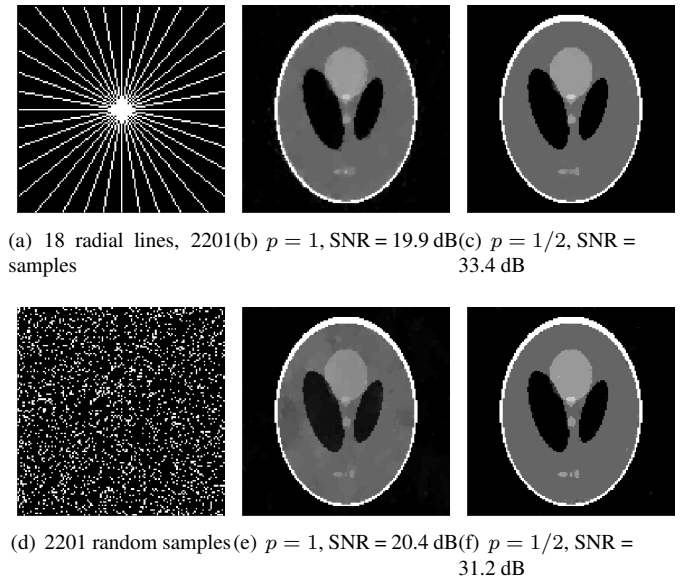
(f) $p = 1/2$, SNR = 31.2 dB

Fig. 2. Now the $p = 1$ case is considered as well. This requires 18 radial lines for the noiseless phantom. Robustness to noise is much poorer than with $p = 1/2$. In both cases, for the noisy phantom random samples give a slightly better visual result than radial sampling, though this is not borne out by the SNR.

[4] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

(a) Ellipse, $\|\nabla x\|_0 = 247$

(b) 3 radial lines, 253 samples

(c) Near reconstruction

(d) 253 random samples

(e) Failed reconstruction

(f) 4 radial lines, 505 samples

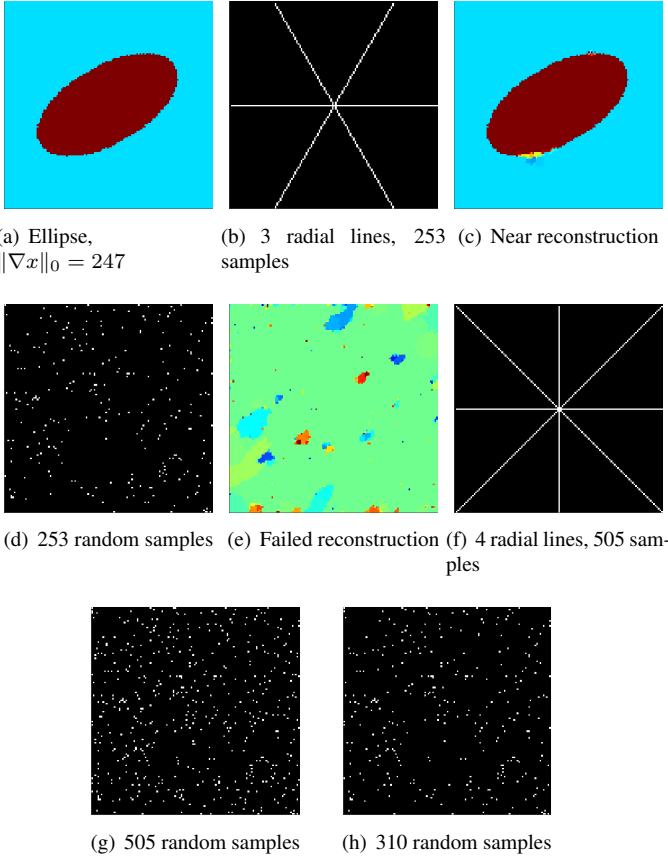(g) 505 random samples

(h) 310 random samples

Fig. 3. A pixellized ellipse is considered. Three radial lines are almost enough, while the same number of random samples fails completely. Four radial lines give perfect recovery, with many fewer random samples being required.

(a)

(b)

Fig. 4. Although random sampling allows exact reconstruction with fewer samples, in the radial case convergence is much faster. For the phantom, 1241 random samples will give a visually exact result (and the same output image in MAT-LAB) sooner, but for the ellipse convergence to any degree is faster with radial sampling. Using the minimal number of random samples is far slower.

[5] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[6] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing." Preprint.

[7] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-ray Sci. Tech.*, vol. 14, pp. 119–139, 2006.

[8] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[9] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.
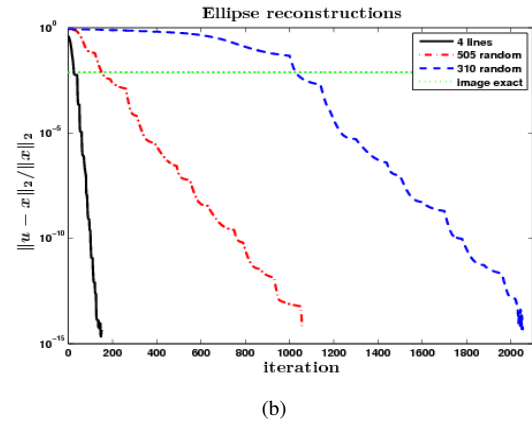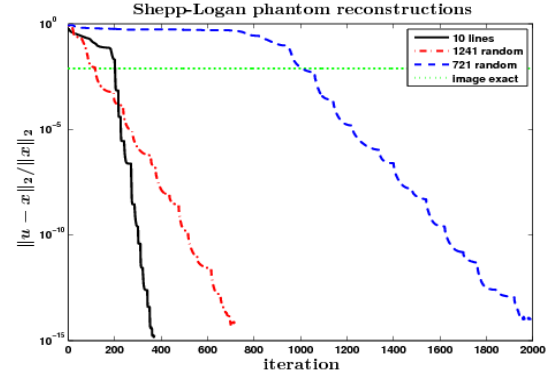
[10] R. Saab, R. Chartrand, and Özgür Yilmaz, "Stable sparse approximations via nonconvex optimization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[11] R. Chartrand, "Nonconvex regularization for shape preservation," in *IEEE International Conference on Image Processing*, 2007.

[12] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

[13] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 187–200, 1999.

[14] R. J. Gardner, "Sets determined by finitely many X-rays," *Geom. Dedicata*, vol. 43, pp. 1–16, 1992.